Relationship between atmospheric CO_2 variations and a satellite-derived vegetation index

C. J. Tucker*, I. Y. Fung†, C. D. Keeling‡ & R. H. Gammon§

* NASA/Goddard Space Flight Center, Code 623, Greenbelt, Maryland 20771, USA
† NASA/Goddard Institute for Space Studies, New York, New York 10025, USA and Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York, New York 10964, USA
‡ Scripps Institution of Oceanography, La Jolla, California 92093, USA
§ NOAA/GMCC, Boulder, Colorado 80302, USA

Estimates derived from satellite data of photosynthetically active radiation absorbed by terrestrial vegetation correlate with atmospheric carbon dioxide concentrations measured at surface recording stations, suggesting that satellite data can be used to estimate terrestrial photosynthesis.

TERRESTRIAL photosynthesis is inversely related to atmospheric carbon dioxide concentrations and is an important component of the global carbon cycle. The lack of temporal resolution and global data has prevented comparisons between estimates of terrestrial photosynthesis and measurements of atmospheric CO_2 concentrations from remote maritime monitoring stations. This is now possible using data derived from meteorological satellites.

Primary production

The biosphere absorbs CO_2 from the atmosphere through photosynthesis which incorporates absorbed carbon into carbohydrates, giving off O_2. In the terrestrial biosphere, CO_2 absorption occurs in the green leaves of plants, where intercepted photosynthetically active radiation (IPAR) (0.4–0.7 μm) is absorbed by the photosynthetic pigments. The energy of the absorbed photons is used to drive the reaction of photosynthesis, yielding carbohydrates through primary production. Incident radiation in the 0.4–0.7-μm region is strongly absorbed by green leaves (by scattering and absorption by photosynthetic pigments), and radiation in the 0.7–1.1-μm region is strongly backscattered or reflected (by scattering in the absence of absorption)[1–5]. Thus, reflected radiance from the red and near-infrared spectral regions can be used to estimate the green-leaf biomass of plant canopies[6,7], are directly be related to the IPAR[8–13], the photosynthetic capacity and the resistance to water-vapour transfer of plant canopies[13].

Atmospheric CO_2

Measurements of atmospheric CO_2 concentrations at monitoring stations have a seasonal oscillation superimposed on an annually increasing trend. The annual amplitude of this oscillation is ~15 parts per 10^6 (p.p.m.) at Point Barrow, Alaska, and decreases southward to ~6 p.p.m. at Mauna Loa, Hawaii, and to ~1–2 p.p.m. in the Southern Hemisphere. The annual maximum atmospheric CO_2 concentration in the Northern Hemisphere occurs in the late spring or early summer and the annual minimum concentration occurs in the autumn[14,15]. The annual oscillation of atmospheric CO_2 is due primarily to seasonal exchange of CO_2 with the terrestrial biosphere[14,16,17]; in contrast, there is little seasonal exchange of CO_2 with the upper ocean[18,19]. The CO_2 exchange with the biosphere has been simulated using diffusion, advection–diffusion, three-dimensional tracer models[20–22] and biospheric-exchange functions. Bolin and Keeling[23] and Pearman and Hyson[24] derived CO_2 sources and sinks consistent with atmospheric CO_2 observations, while Azevedo[25] constructed simple biospheric-exchange functions based on assumed vegetation seasonality and geographical location. However, the biospheric CO_2 drawdown and release functions that correlated best with measured CO_2 concentrations showed differences by at least a factor of 2 (refs 20–25). The release of CO_2 from terrestrial vegetation can be approximated if primary production and the temperature regime are known; the drawdown of CO_2 by terrestrial vegetation can be determined if primary production is known. Thus, elucidation of the role of the terrestrial biota in the gobal carbon cycle requires an estimate of the terrestrial primary production; this can be achieved through remote sensing by satellite.

Remote sensing

By monitoring global changes in the IPAR of the terrestrial biota, an estimate of CO_2 drawdown and primary production can be obtained through the relationships between leaf structure, leaf density and light in plant canopies, and that between IPAR and photosynthetic capacity[13].

The CO_2 cycle involves not only photosynthesis and respiration, but also decomposition of carbon compounds, uptake and release by oceans, combustion of fossil fuels, and fires involving vegetation. Deforestation contributes towards atmospheric CO_2 through burning and/or subsequent decomposition of phytomass, litter and associated carbon concentrations in the soil, although the magnitude of this contribution is uncertain[26,27]. Remote sensing by satellite enables the temporal and spatial variations in the IPAR of the terrestrial biota to be quantified.

Remote-sensing studies of the biota have used Landsat and meteorological satellites data. Landsat multispectral scanner (MSS) data, with a spatial resolution of 80 m, are not appropriate for the study of seasonal changes in the terrestrial biota, because ~5,000 Landat MSS scenes would be required to cover the world's surface. The polar-orbiting meteorological satellites of the National Oceanic and Atmospheric Administration (NOAA) are suitable for observing terrestrial IPAR dynamics. The advanced very-high-resolution radiometer (AVHRR) sensor on board these polar-orbiting satellites daily collects global radiance data at a 4-km spatial resolution. Combinations of channel 1 (0.55–0.68 μm) and channel 2 (0.73–1.1 μm) data from the NOAA satellites' AVHRR have been used to study continental land cover, global vegetation phenology and net primary production by grasslands using data from 1 to 4-yr periods[28–31].

Data methods and use

Data on monthly atmospheric CO_2 concentrations for 1982–84 were obtained for Point Barrow, Alaska (147° W, 71° N), and

§ Present address: NOAA/PMEL, Seattle, Washington 98115, USA.

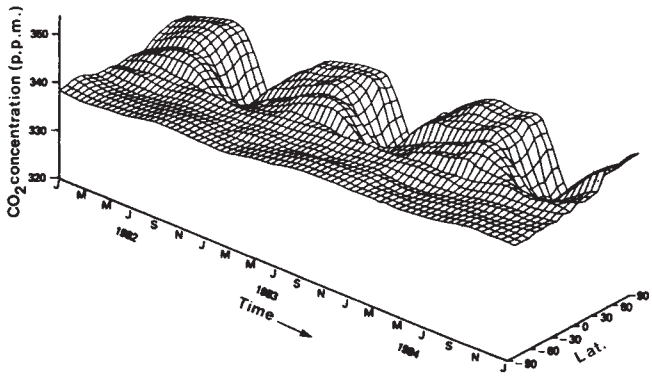

Fig. 1 Variation of global atmospheric CO_2 concentrations with latitude and time based on the NOAA/GMCC flask measurements for 1982–84.

Mauna Loa, Hawaii (156° W, 20° N); and from January 1982 to September 1984 for the South Pole (90° S). Stations there are operated by the NOAA Geophysical Monitoring for Climatic Change (GMCC) programme, and the CO_2 flask samples are analysed at the Scripps Institute of Oceanography. CO_2 concentrations for 20 globally dispersed monitoring sites operated and analysed by NOAA/GMCC for 1982–84 were also obtained.

The atmospheric CO_2 concentration is the combined response to terrestrial biospheric, oceanic and anthropogenic sources and sinks, convolved with long-range transport. Figure 1 shows the total CO_2 signal from January 1982 to December 1984 measured at the 20 global monitoring stations. An annually increasing trend of ~1.5 p.p.m. yr^{-1} from anthropogenic sources, and an annual cycle, with an amplitude greatest around 50°–70° N, is evident. To extract the seasonal variation from the CO_2 records, we computed the 12-month running means for each station and subtracted these from the CO_2 record (see ref. 15).

Global weekly data from NOAA-7's AVHRR were obtained for the period from 12 April 1982 to 28 October 1984 (ref. 32). These data were re-mapped from a polar stereographic projection to a Mercator projection so that the world could be represented in a single image. The normalized difference vegetation index (NDVI) $[(C_2-C_1)/(C_2+C_1)]$ was calculated[6,7], and the highest monthly NDVI value was selected from the weekly data for each month between April 1982 and October 1984 (Table 1). This procedure minimizes the effects of Sun angle, off-nadir viewing, atmospheric pathlength and aerosols, and clouds, all of which decrease the NDVI[33]. In winter conditions, spurious NDVI values result from a combination of low solar illumination levels, differing dark-current levels in channels 1 and 2, and pathlength-associated variation in atmospheric effects. These effects are apparent not only during the respective hemisphere winters, but also if the NDVI is calculated from AVHRR data acquired during the night. To compensate for these spurious values, mean monthly surface temperature data were acquired[34], and areas within the 0 °C or colder regions were assigned NDVI values of 0.00 because physiological evidence suggests that the rate of photosynthesis at 0 °C is close to zero[35].

The monthly NDVI data, which have an average grid-cell size of ~40 km, were used to compute average values for global latitude zones. The zonal averages were multiplied by the land area within each zone (to weight the averages by the land area involved) and summed for various latitude zones; the sum for each larger zone was then divided by the total land area in that zone. The NDVI for a latitude zone represents the potential IPAR; hence, we have assumed that the NDVI is directly related

Table 1 Weekly data used for monthly NDVI composite images

Month	Year	Weeks used	Comments
April	1982	12–18, 19–25, 26–30 April; 1–2 May	Includes 1–2 May
May	1982	3–9, 10–16, 17–23, 24–30 May	
June	1982	31 May; 1–6, 7–13, 14–20, 21–27 June	Includes 31 May
July	1982	5–11, 12–18, 19–25, 26–31 July; 1 August	Includes 1 August
August	1982	2–8, 9–15, 16–22, 23–29 August	
September	1982	30–31 August; 1–5, 6–12, 13–19, 20–26 September	Includes 30 and 31 August
October	1982	4–10, 11–17, 18–24, 25–31 October	
November	1982	1–7, 8–14, 15–21, 22–28 November	
December	1982	29–30 November; 1–5, 6–12, 13–19, 20–26 December	Includes 29 and 30 November
January	1983	3–9, 10–16, 17–23, 24–30 January	
February	1983	31 January; 1–6, 7–13, 14–20, 21–27 February	Includes 31 January
March	1983	28 February; 1–6, 7–13, 14–20, 21–27 March	Includes 28 February
April	1983	4–10, 11–17, 18–24, 25–30 April; 1 May	Includes 1 May
May	1983	2–8, 9–15, 16–22, 23–29 May	
June	1983	30–31 May; 1–5, 6–12, 13–19, 20–26 June	Includes 30 and 31 May
July	1983	4–10, 11–17, 18–24, 25–31 July	
August	1983	8–14, 15–21, 22–28 August	No data 1–7 August
September	1983	5–11, 12–18, 19–25, 26–30 September; 1–2 October	Includes 1 and 2 October
October	1983	3–9, 10–16, 17–23, 24–30 October	
November	1983	31 October; 1–6, 7–13, 14–20, 21–27 November	Includes 31 October
December	1983	5–11, 12–18, 19–25, 26–31 December; 1 January	Includes 1 January
January	1984	2–8, 9–12, 16–22, 23–29 January	No data 13–15 January
February	1984	30–31 January; 1–5, 6–12, 27–28 February; 1–4 March	Includes 30–31 January and 1–4 March; no data 13–26 February
March	1984	5–11, 12–14, 19–25 March	No data 14–18, 26–31 March
April	1984	2–8, 9–14, 16–21, 23–29 April	No data 15 and 22 April
May	1984	30 April; 1–6, 7–13, 14–20, 21–27 May	Includes 30 April
June	1984	4–10, 11–17, 18–24, 25–30 June; 1 July	Includes 1 July
July	1984	2–8, 9–15, 16–22, 23–29 July	
August	1984	30–31 July; 1–5, 6–12, 13–19, 20–26 August	Includes 30 and 31 July
September	1984	3–9, 10–16, 17–23, 24–30 September	
October	1984	1–7, 8–14, 15–21, 22–28 October	

The maximum monthly value was selected from the weekly data.

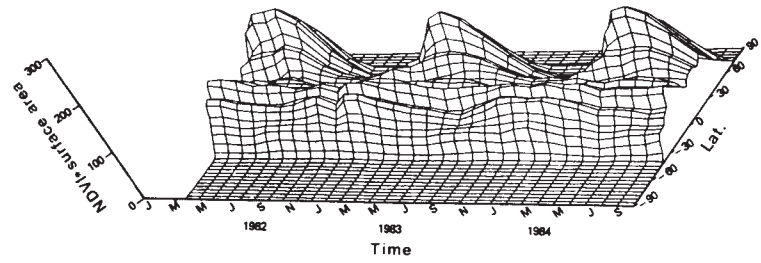

Fig. 2 Weighted NDVI data plotted against time and latitude zone. Note the highly seasonal effects in the northern latitudes, the influence of deserts in the 20°–30° N latitude zone, the generally constant response in equatorial areas, and the influence of the low proportion of land area south of 30° S. See also Fig. 1. The average NDVI value for each 5° latitude zone has been multiplied by the land area in the zone to compensate for global variations in land area.

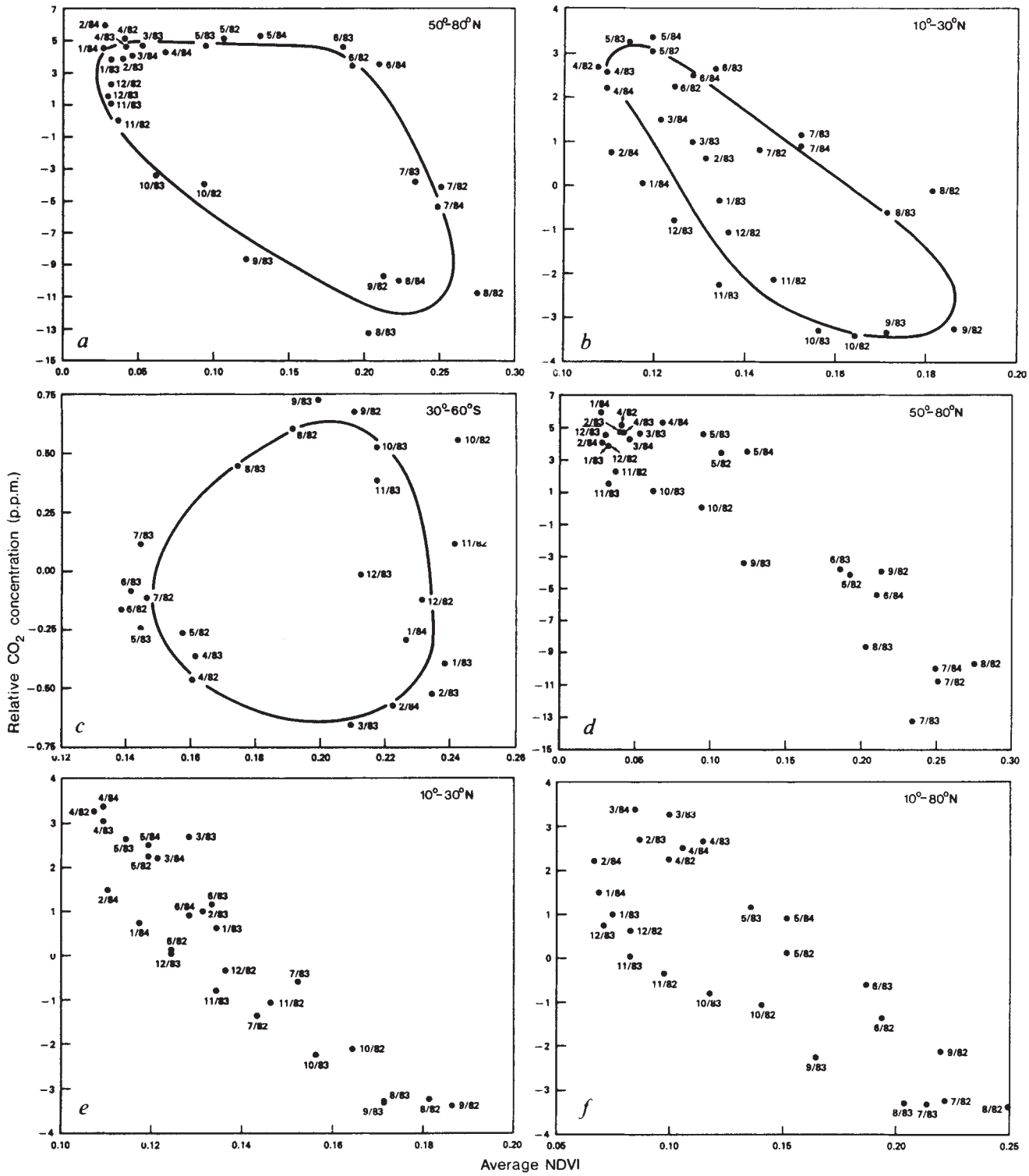

Fig. 3 Relationships between atmospheric CO_2 concentrations and NDVI averages with and without time lags between NDVI measurements and CO_2 measurements. The CO_2 data are from stations at Point Barrow, Alaska: *a*, no time lag; *d*, 1-month lag; Mauna Loa, Hawaii: *b*, no time lag; *e*, 1-month lag; *f*, 2-month lag; South Pole: *c*, no time lag. The corresponding NDVI averages are from 50° to 80° N for Point Barrow, 10° to 30° N and 10° to 80° N for Mauna Loa, and 30° to 60° S for the South Pole.

to the photosynthetic capacity[13] of the terrestrial surface in that zone and should be inversely related to atmospheric CO_2 concentrations. Figure 2 shows the latitudinal and temporal variation of the zonally averaged NDVI (weighted by land area) from April 1982 to October 1984. Consistent with considerations of growing seasons, the maximum zonally averaged and weighted NDVI occurs in summer at middle to high latitudes in the Northern Hemisphere. This pattern is a consequence both of the large proportion of land area and the highly seasonal behaviour of vegetation at these latitudes. The Southern Hemisphere south of 30° S has a low weighted NDVI because its vegetation is not as seasonal and the proportion of land area is less than in latitude zones north of 30° S. The weighted NDVI in equatorial areas is high throughout the year, and a desert belt is apparent as a depressed weighted NDVI from ~20° N to ~30° N. Figure 2 also shows differences in the weighted NDVI from 1982 to 1984. The 1982 global peak occurred in August, whereas the 1983 and 1984 global peaks occurred in July.

Results

The data presented in Figs 1 and 2 reveal an inverse relationship between the NDVI and the seasonal variations in atmospheric CO_2: in the Northern Hemisphere, the NDVI increased in spring to summer as the atmospheric CO_2 concentration decreased.

The NDVI data were plotted against the atmospheric CO_2 data from the Point Barrow, Mauna Loa and South Pole stations (Fig. 3). The South Pole data (Fig. 3*c*) showed no relationship between NDVI and CO_2: south of 30° S, the oceanic area is greater than land area which leads to smaller land/sea contrasts in air temperature. Consequently, seasonal behaviour of both the NDVI and the atmospheric CO_2 concentrations is less marked than in corresponding latitudes of the Northern Hemisphere (Figs 1, 2). Remote sources of CO_2 may also dominate or obscure local sources of CO_2 through atmospheric circulation and mixing in this region.

In contrast, the CO_2 data from Point Barrow and Mauna Loa and the NDVI data from 50 to 80° N and from 10 to 30° N, respectively (Figs 3*a*, *b*), showed a circular/elliptical relationship. At Point Barrow, the data showed no change in relative CO_2 concentration from April to June of 1982, 1983 and 1984, while the NDVI from 50 to 80° N increased (Fig. 3*a*). At Mauna Loa, however, the data showed a decrease in relative CO_2 concentration from May to June of 1982 and 1983, while the NDVI from 10 to 30° N again increased (Fig. 3*b*).

The Point Barrow results are consistent with heterotrophic respiration rates reported for soil and litter in the tundra and taiga zones between April and June[36,37]. As soon as the ground thaws, soil and litter respiration releases a burst of CO_2 and this could explain the April to June displacement of the relative CO_2–NDVI relationship in Fig. 3*a*.

Atmospheric circulation and boundary-level mixing effects would be expected to delay equilibration of CO_2 concentrations by several weeks. A time lag between the NDVI and the measured atmospheric CO_2 concentrations would be expected because of the distance between the Point Barrow and Mauna Loa stations and mid-latitude forest sites which have maximum seasonality in photosynthesis. When the NDVI for month n was plotted against the atmospheric CO_2 concentration for month $n+1$, approximately linear relationships were found with the Point Barrow and Mauna Loa data (Fig. 3*d*, *e*). The negative slopes in Fig. 3*d* and *e* indicate the increased absorption of CO_2 by increased amounts of terrestrial photosynthetic activity.

Inverse linear relationships, found between the Mauna Loa CO_2 concentrations, lagged by 2 months, and the NDVI for 10–80° N (Fig. 3*f*) showed that CO_2 oscillations at Mauna Loa represented the CO_2 response to the seasonal vegetation dynamics of the Northern Hemisphere and were not dominated by proximate sources and sinks. In addition, the globally averaged CO_2 cycle has a peak-to-peak amplitude of 4–5 p.p.m., about 75% of that at Mauna Loa, and is nearly in phase with the cycle there. The annual cycle at Mauna Loa is thus characteristic of the globally averaged cycle.

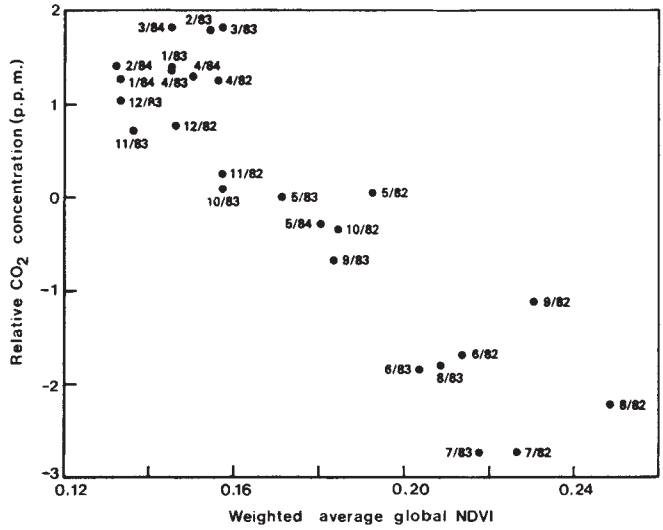

Fig. 4 The globally averaged atmospheric CO_2 concentration plotted against the globally averaged NDVI with a time lag of 1 month. The CO_2 data are from the global network of 20 NOAA/GMCC stations.

Figure 4 shows the globally averaged CO_2 measurements from the 20 stations plotted against the globally averaged NDVI. When a lag of 1 month was introduced into the CO_2 data, an inverse linear relationship was found, suggesting that the global NDVI represents the photosynthetic capacity of the terrestrial biosphere.

Conclusions

Our data show the relationship between monthly variations in atmospheric CO_2 concentrations and terrestrial NDVI dynamics, which are, in turn, estimates of IPAR and photosynthetic capacity. Note that we have not considered factors such as respiration, decomposition, consumption of fossil fuels, oceanic processes and burning of vegetation, all of which influence atmospheric CO_2 concentrations. Our analysis demonstrates the measurable link between atmospheric CO_2 drawdown and terrestrial NDVI dynamics and suggests that there may be quantitative relationships between multi-temporal satellite data and atmospheric CO_2 drawdown. The inclusion of oceanic CO_2 uptake and release, respiration and decomposition processes, fossil fuel contributions and other CO_2 sources and sinks can be used to advance our quantitative understanding of the global CO_2 cycle.

We thank T. Goff, B. Rank, E. Matthews, K. Prentice and J. Lerner for assistance, T. J. Conway for the 1984 NOAA/GMCC flask data, and A. M. Heasty for comments. This research was supported in part by NASA Life Sciences Division, by NASA Earth Observation Division, and by NASA cooperative agreement NCC5-29 with Columbia University, as well as by the Carbon Dioxide Research Division, Office of Energy Research, US Department of Energy, under contract 19X-2229C with the Lamont-Doherty Geological Observatory of Columbia University with Martin Marietta Energy Systems, Inc., under contract DE-AC05-840R21400.

Received 6 February; accepted 31 October 1985.

1. Gates, D. M., Keegan, H. J., Schleter, J. C. & Weidner, V. P. *Appl. Opt.* **4**, 11–20 (1965).
2. Knipling, E. B. *Remote Sens. Envir.* **1**, 115–119 (1970).
3. Woolley, J. T. *Pl. Physiol.* **47**, 656–662 (1971).
4. Kumar, R. & Silva, L. *Appl. Opt.* **12**, 2950–2954 (1973).
5. Tucker, C. J. & Garratt, M. W. *Appl. Opt.* **16**, 635–642 (1977).
6. Tucker, C. J. *Grass Forage Sci.* **35**, 177–182 (1980).
7. Curran, P. J. *Phil. Trans. R. Soc.* A**309**, 257–270 (1983).

8. Kumar, M. & Monteith, J. L. in *Plants and the Daylight Spectrum* (ed. Smith, H.) 133-144 (Academic, London, 1982).
9. Daughtry, C. S. T., Galio, K. P. & Bauer, M. E. *Agron. J.* **75**, 527-531 (1983).
10. Hatfield, J. L., Asrar, G. & Kanemasu, E. T. *Remote Sens. Envir.* **14**, 65-76 (1984).
11. Wiegand, C. L. & Richardson, A. J. *Agron. J.* **76**, 543-548 (1984).
12. Asrar, G., Fuchs, M., Kanemasu, E. T. & Hatfield, J. L. *Agron. J.* **76**, 300-306 (1984).
13. Sellers, P. J. *Int. J. Remote Sens.* **6**, 1335-1372 (1985).
14. Keeling, C. D. *Proc. Carbon Dioxide Research Conference: Carbon Dioxide, Science and Consensus* 2.1-2.62 (Department of Energy Conf.-82097, 1982).
15. Komhyr, W. D. *et al. J. geophys. Res.* **90**, 5567-5596 (1985).
16. Machta, L., Hansen, K. & Keeling, C. D. in *The Fate of Fossil Fuel CO_2 in the Oceans* (eds Anderson, N. R. & Malahoff, A.) 131-144 (Plenum, New York, 1977).
17. Woodwell, G. M., Houghton, R. A. & Tempel, N. R. *J. geophys. Res.* **78**, 932-940 (1973).
18. Fung, I. Y. in *Proc. 6th Oak Ridge natn. Lab. Life Sci. Symp. on the Global Carbon Cycle* (ed. Jacobs, V.) (US Department of Energy, Oak Ridge, in the press).
19. Heimann, M. M., Keeling, C. D. & Fung, I. Y. in *Proc. 6th Oak Ridge natn. Lab. Life Sci. Symp. on the Global Carbon Cycle* (ed. Jacobs, V.) (US Department of Energy, Oak Ridge, in the press).
20. Junge, C. E. & Czeplak, G. *Tellus* **20**, 422-434 (1968).
21. Machta, L. *Bull. Am. met. Soc.* **53**, 402-420 (1972).
22. Fung, I. Y., Prentice, K., Mathews, E., Lerner, J. & Russel, G. *J. geophys. Res.* **88**, 1281-1294 (1983).
23. Bolin, B. & Keeling, C. D. *J. geophys. Res.* **68**, 3899-3920 (1963).
24. Pearman, G. I. & Hyson, P. *J. geophys. Res.* **85**, 4457-4467 (1980).
25. Azevedo, A. E. thesis, Columbia University (1982).
26. Woodwell, G. M. *et al. Science* **222**, 1081-1086 (1984).
27. Woodwell, G. M. in *The Role of Terrestrial Vegetation in the Global Carbon Cycle: Measurement by Remote Sensing* (ed. Woodwell, G. M.) 1-17 (SCOPE 23, Wiley, New York, 1984).
28. Tucker, C. J., Townshend, J. R. G. & Goff, T. E. *Science* **227**, 369-375 (1985).
29. Justice, C. O., Townshend, J. R. G., Holben, B. N. & Tucker, C. J. *Int. J. Remote Sens.* **6**, 1278-1318 (1985).
30. Goward, S. N., Tucker, C. J. & Dye, D. G. *Vegetacio* (in the press).
31. Tucker, C. J., Vanpraet, C. L., Sharman, M. J. & Van Ittersum, G. *Remote Sens. Envir.* **17**, 233-249 (1985).
32. Tarpley, J. D., Schneider, S. R. & Money, R. L. *J. Climate appl. Met.* **23**, 491-493 (1984).
33. Holben, B. N. & Fraser, R. S. *Int. J. Remote Sens.* **5**, 145-160 (1984).
34. Schutz, C. & Gates, W. L. Preprint (Rand Corporation, R-1029-ARPA, Santa Monica, 1972).
35. Larcher, W. & Bauer, H. in *Physiological Plant Ecology* (eds Lange, O. L., Nobel, P. S., Osmond, C. B. & Ziegler, H.) 403-437 (Springer, Berlin, 1981).
36. Swift, M. J., Heal, O. W. & Anderson, J. M. *Decomposition in Terrestrial Ecosystems*, 308-309 (Blackwell, Oxford, 1979).
37. Linkins, A. E., Melillo, J. M. & Sinsabaugh, R. L. *Current Perspectives in Microbial Ecology* (eds Klug, M. J. & Reddy, C. A.) 572-579 (American Society of Microbiology, Washington, DC, 1984).

Solvation energy in protein folding and binding

David Eisenberg & Andrew D. McLachlan*

Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90024, USA
* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

We have developed a method for calculating the stability in water of protein structures, starting from their atomic coordinates. The contribution of each protein atom to the solvation free energy is estimated as the product of the accessibility of the atom to solvent and its atomic solvation parameter. Applications of the method include estimates of the relative stability of different protein conformations, estimates of the free energy of binding of ligands to proteins and atomic-level descriptions of hydrophobicity and amphiphilicity.

OF the forces that guide a polypeptide chain to its folded form in water, solvent interactions, including the hydrophobic interaction, are thought to be among the most important[1-4]. Yet these interactions are currently evaluated by relatively primitive methods. This is in contrast to the other component energies that stabilize proteins, such as hydrogen bonds, for which atom-atom potential functions have been devised[5-7]. Present methods for estimating the contribution of solvation energy to protein stability include the assumption that the hydrophobic character of each amino-acid residue can be summarized by a single number, the amino-acid hydrophobicity[8-12]. This is an oversimplification for amino acids such as Trp, Tyr, Glu, Gln, Lys and Arg which have both polar and apolar parts. Some authors have tried to overcome this limitation by devising more elaborate residue hydrophobicities to improve the description[8,13,14]. Also an oversimplification is the commonly used approximation that the hydrophobic energy of a folded protein molecule is proportional to the total protein surface that is accessible to water, regardless of whether the exposed surface is apolar, polar or charged[15].

To develop an explicitly atomic description of the interaction of water with a protein, we extend the ideas of Langmuir[16], Cohn and Edsall[17], and others[18,19]. The basic assumption is that the free energy of interaction of a solute with water can be considered as a sum of energies of atomic groups. We follow Langmuir in using the exposed surface areas of a group as a measure of its interaction with solvent. In practice we use the solvent-accessible surface area of Lee and Richards[20] and other recent workers[21-24]. This is defined as the area over which the centre of a water molecule of radius 1.4 Å can move while maintaining unobstructed contact with the group. In our method, the sign and strength of the water-solvent interaction are specified by the atomic solvation parameter (ASP) of each atom accessible to water. These values are determined, as described below, from free energies of transfer[8-12]. Thus, our method effectively combines two common approaches for evaluating hydrophobic forces: computation of solvent-accessible surface areas; and estimating energies from free energies of transfer. Our method, however, extends the first approach by weighting the effect of each atom by its polar or apolar character, and extends the second in permitting calculation of the solvation energy from the coordinates of individual atoms in each residue. It advances both methods in permitting estimates of the free energy of transfer of small molecules and of the contribution of solvation to the free energy of binding of small molecules to proteins.

Atomic contributions

Here we express the contribution of protein-solvent interactions to the free energy of protein folding as a sum over all atoms of the structure (except hydrogen atoms, which are not treated explicitly). The term for each atom is the product of its solvent-accessible area[20] and its ASP for transfer from the interior of a protein to aqueous solution. The transfer to solution of a single atom, i, immersed in protein is depicted in Fig. 1*a*. Let the accessible surface area of the atom be given by A_i, and its atomic solvation parameter be given by $\Delta\sigma_i$. Then the free energy of transfer (ΔG) is

$$\Delta G_i = \Delta\sigma_i A_i \quad (1)$$

Now consider the transfer to water of amino-acid residue R immersed in protein (Fig. 1*a*). We assume that the change in free energy for this process can be approximated as a sum of atomic terms, each like that of equation (2)

$$\Delta G_R = \sum_{\text{atoms i}} \Delta\sigma_i A_i \quad (2)$$

Note that the areas A_i now depend on the conformation of the amino acid, and that some atoms will be obscured by their